Data Privacy and Health Research

Harry A. Guess, MD, PhD
Vice President, Epidemiology
Merck Research Laboratories
Merck & Co., Inc.

2003-ms-1121
Outline

• Health Research: Clinical and Epidemiologic Research
• Human subjects protection in health research
• Identifiable data & privacy protection
  – Informed consent: feasibility & impact on validity
  – De-identification: feasibility & impact on validity
  – Statistical certification that identification risk is “very small”
  – Limited Use Data Sets and Data Use Agreements
• Public health importance of studies using identifiable medical data
• Conclusions
Clinical and Epidemiologic Research: Two different paradigms

**Clinical Research**
- Typically prospective
- Typically involves at most a few thousand patients, except in the most expensive of studies
- Research risks may involve possibility of physical harm (e.g., from adverse reactions)
- Study-specific informed consent is easily obtained as part of the investigator-patient interaction

**Epidemiologic Research**
- Often retrospective
- Often involves analyses of medical records that have been collected previously and for other purposes from many thousands or millions of patients
- Research risks have to do with potential harm from release of health information
- Study-specific informed consent is often impossible to obtain without either invalidating the study or making it prohibitively expensive
Human Subjects Protection in Epidemiologic Research

• Ethical principles* are the same as for clinical research
  – Respect for persons
    • Underlies the requirement for oversight of research by an Institutional Review Board (IRB) / Privacy Board / Ethics Committee
    • Requires patient informed consent - with certain well-defined exceptions
  – Beneficence
    • Requires that risks of the research be reasonable in relation to possible benefits and that any risks to research subjects be minimized
  – Justice
    • Requires fairness in sharing risks and benefits of research


See NIH Ethics Training Website:  http://ohsr.od.nih.gov/extramural/extramural_training.html
Conditions in “The Common Rule” (45 CFR Part 46) under which the requirement of obtaining informed consent for human subjects research may be waived

• An Institutional Review Board (IRB) may alter or waive the requirements to obtain informed consent if it finds and documents that:
  – (1) The research involves no more than minimal risk to the subjects;
  – (2) The waiver or alteration will not adversely affect the rights and welfare of the subjects;
  – (3) The research could not practicably be carried out without the waiver or alteration; and
  – (4) Whenever appropriate, the subjects will be provided with additional pertinent information after participation.

45 CFR 46.116(d)  

CFR = U.S. Code of Federal Regulations
Health Insurance Portability and Accountability Act of 1996
HIPAA

• **45 CFR 160-164**: Standards for Privacy of Individually Identifiable Health Information

• This regulation is the second final regulation to be issued in the package of rules mandated under title II subtitle F section 261–264 of the Health Insurance Portability and Accountability Act of 1996 (HIPAA), Public Law 104–191, titled “Administrative Simplification.”
Privacy Protections for Human Research Participants

• The HIPAA Privacy Prohibition

• The HIPAA Privacy Standards generally state that a “covered entity may not use or disclose protected health information (PHI), except as permitted or required by” the regulation.

• 45 CFR 164.502
HIPAA: Some Definitions

Covered Entity
Health plan; Health care clearinghouse; Health care provider that transmits any health information in electronic form.

Protected Health Information (PHI)
Individually Identifiable Health Information (IIHI) in the possession of a Covered Entity, whether transmitted or maintained through electronic media, in hard copy, or by other means.

Individually Identifiable Health Information (IIHI)
Information about the physical or mental health of an individual that is created or received by a covered entity and that identifies or can reasonably be used to identify the individual.
# HIPAA: Privacy Protections for Human Research Participants

## The Common Rule

<table>
<thead>
<tr>
<th>Applies to <em>federally funded or regulated</em> research</th>
<th>Applies to <em>all</em> research for which HIPAA covered entities use or disclose PHI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protects <em>rights</em> and welfare</td>
<td>Protects <em>privacy</em> and welfare</td>
</tr>
<tr>
<td><em>Human subject</em>: a living individual; subject of data or information</td>
<td><em>Individual</em>: subject of information; a living or deceased person</td>
</tr>
<tr>
<td>Establishes IRBs</td>
<td>Establishes <em>Privacy Boards</em>; recognizes IRBs</td>
</tr>
<tr>
<td>Board reviews all research protocols</td>
<td>Board reviews only the waiver or alteration of authorization (some unresolved issues on scope of required review of consents)</td>
</tr>
<tr>
<td>Annual and continuing reviews</td>
<td><em>No requirement</em></td>
</tr>
<tr>
<td>Informed Consent</td>
<td>Authorization and Consent</td>
</tr>
</tbody>
</table>
HIPAA: Effect on Health Researchers

- HIPAA applies to covered entities, not covered individuals.
  - Covered entities include healthcare providers that communicate electronically with health plans or with other providers for the purpose of sending or receiving referrals required by health plans.
- Thus, a health researcher, who is not a healthcare provider or other covered entity, who receives PHI from a covered entity is not directly subject to the HIPAA regulation.
- Indirect control on how a researcher can access data from a covered entity is provided by HIPAA in several ways, including
  - through requiring that the covered entity obtain authorization from each patient whose data will be used for anything other than treatment, payment, or healthcare operations
  - or through detailed criteria that an IRB or Privacy Board must find to have been met in order for a covered entity to release PHI for a research study without specific authorization from each patient
Patient authorization is not required for research use of PHI when*

• IRB or privacy board waiver of authorization according to specific detailed criteria, or
• using PHI to help prepare a research protocol or to pre-screen patients for research, where no PHI is removed from the covered entity, or
• research using PHI of deceased persons
• only de-identified PHI is used, or
• a “limited data set” is used under a “data use agreement”

*45 CFR 164.512
De-Identification of health information:

What are the requirements in HIPAA?
To “de-identify” medical data under HIPAA, one must either remove all of the following 18 identifiers or provide an expert statistical determination that the risk of identifying an individual would be very small*:

- Names
- All geographic subdivisions smaller than a State (some complex exceptions)
- All elements of dates (except year) for dates ... including birth date, admission date, discharge date, date of death; and all ages over 89...
- Telephone numbers
- Fax numbers
- Electronic mail addresses
- Social security numbers
- Medical record numbers
- Health plan beneficiary numbers
- Account numbers
- Certificate/license numbers
- Vehicle identifiers and serial numbers, including license plate numbers
- Device identifiers and serial numbers
- Web Universal Resource Locators (URLs)
- Internet Protocol (IP) address numbers
- Biometric identifiers, including finger and voice prints
- Full face photographic images and any comparable images and
- Any other unique identifying number, characteristic, or code

*45 CFR 164.514
Removing all 18 identifiers renders such data sets essentially useless for Longitudinal Epidemiologic Research

• In particular, removing all dates makes it impossible to determine the temporal sequence in which events happen for any individual
  – Causes must precede effects, but without dates it is not possible to separate potential causes from potential consequences

• Use of the data to help establish cause and effect is therefore essentially impossible
Expert statistical determination that the risk of identifying an individual is “very small”*

“A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable:

(i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and

(ii) Documents the methods and results of the analysis that justify such determination…."

*45 CFR 164.514
……as a starting point, the Secretary approves the use of the following as guidance to such generally accepted statistical and scientific principles and methods:


The Checklist on Disclosure Potential of Proposed Data Releases (http://www.fcsm.gov/docs/checklist/799.doc) (prepared by the Confidentiality and Data Access Committee, Federal Committee on Statistical Methodology, Office of Management and Budget)


Comment: I downloaded the first; the second appears no longer to be available
August 2002 HIPAA Final Rule

- The August 2002 Final Rule contained a number of changes affecting use of protected health information for research
- One change facilitating epidemiologic research is the ability to obtain a “limited use data set” under a “data use agreement” without individual patient authorization*

*45 CFR 164.514
To create a “limited data set” under HIPAA, one must remove all of the following 16 identifiers and sign a “data use agreement” that establishes permitted uses, requires safeguards to prevent disclosure, bars unauthorized further disclosure, and forbids attempts to identify or contact the individuals:

- Names
- Postal address except city, state, 5-digit zip code
- Telephone numbers
- Fax numbers
- Electronic mail addresses
- Social security numbers
- Medical record numbers
- Health plan beneficiary numbers
- Account numbers
- Certificate/license numbers
- Vehicle identifiers and serial numbers, including license plate numbers
- Device identifiers and serial numbers
- Web Universal Resource Locators (URLs)
- Internet Protocol (IP) address numbers
- Biometric identifiers, including finger and voice prints
- Full face photographic images and any comparable images
Limited use data sets should help lessen the adverse impact of HIPAA on epidemiologic research

• Limited data sets **can contain dates**
  – this is essential for longitudinal studies of medical care outcomes

• Health outcomes in very young children (e.g., neonates) can be studied, since the data sets can contain age in years, months, days, and hours if necessary for the research

• Health outcomes in the very elderly can be studied more effectively because the data sets can contain the exact year of age of such patients, if necessary for the research

• Geographic variations in disease patterns can be studied since locations by 5-digit zip codes can be specified
Why such caution in restricting release of protected health information?
Use of external public information in combination with databases can reveal patient names and addresses.

Database information: DOB, Gender, Zip Code, …

Public information: e.g. voter registration rolls linking names and addresses with DOB, gender, Zip Codes

= Identification of the patient names with the medical information in the database
Using public information to uniquely identify people
Cambridge, Massachusetts Voting List - 54,805 voters

• **Data Elements**

  - Birth date alone
  - Birth date and gender
  - Birth date and 5-digit ZIP Code
  - Birth date and full postal code

• Percent of voters whose names and addresses were uniquely identified by the data elements

  - 12%*
  - 29%
  - 69%
  - 97%

Sweeney L: Journal of Law, Medicine & Ethics. 1997; 25:98-110

*These results are not surprising. This is about what one would expect from the approximation \((1 - 1/k)^{N-1} \approx \exp(-N/k)\), where \(k = \) available dates of birth in a voter cohort and \(N = 54,085\). The other numbers look plausible, by similar reasoning.
• What may appear to be “de-identified” patient data may be combined with public data to link patient names with medical information in databases.

• Sweeney’s results appear to have had a considerable influence on HIPAA standards for what constitutes “de-identified” data.

• These results also imply that making a statistical determination of “very small risk” could itself be risky for the statistician and his employer if he fails to know about some obscure but “public” data set which could be used together with the health data to identify an individual.

• Standards remain to be defined for what is an acceptable level of “due diligence” in making such a determination.

• It is likely that IRBs would be reluctant to accept such a statistical determination until standards are better defined.
Some states have required individual patient informed consent for all medical records research

- In 1996, Minnesota enacted a law that placed stringent consent requirements on the use of patient data for research.

- Records created since January 1, 1997 could not be used for research without the patient’s written authorization.
Effect of Minnesota legislation requiring specific informed consent on response rate for a medical records study

• STUDY DESIGN: Seizures associated with a pain medication -- part of FDA post-marketing surveillance to evaluate adverse events with approved drugs.

• DATA COLLECTION: Informed consent for Minnesota plan members consisted of: (1) letter from health plan medical director, (2) 2nd mailing to non-respondents, and (3) a follow-up telephone call to non-respondents.

• PRINCIPAL FINDING --- very low participation rates:
  – 19% (26/140) of health plan members in Minnesota, where informed consent was required, returned a signed consent form
  – In 5 other states, where patient informed consent was not required, health care providers granted access to patient medical records for 93% (123/132) of the members.

• CONCLUSION: Legislation requiring study-specific consent was associated with low participation and increased time to completion. Efforts to protect privacy may conflict with ability to produce valid research to safeguard and improve public health.

• The Minnesota law has subsequently been amended to permit use of records where the patient does not respond to 2 requests for authorization mailed to the patient’s last known address.

• At Mayo Clinic, that change decreased the percentage of patient records that the patient consent requirement made unavailable for studies from 20.7 percent to 3.2 percent. Mayo Clinic researchers remain concerned that variations in the rate of refusal among different patient groups, for example, young versus old, may tend to skew the results obtained from these data\textsuperscript{1,2}.

\textsuperscript{1}MEDICAL PRIVACY REGULATION. GAO Report 01-584. April 2001.
\textsuperscript{2}S. J. Jacobsen and others, “Potential Effect of Authorization Bias on Medical Record Research,” Mayo Clinic Proceedings, Vol. 74, No. 3 (April 1999), p. 333
Why medical records studies are important in public health

- To monitor the health of populations and to detect emerging disease problems, e.g., trends and patterns in asthma, renal disease, coronary heart disease, cancer.
- To identify populations at high risk for disease and to identify factors that are either potentially harmful or helpful.
- To determine the effectiveness of health interventions as they are used in clinical practice, e.g., monitoring effects of new vaccines.
- To quantify prognosis, e.g., survival statistics for various stages and grades of cancer and for cardiovascular disease.
- To assess usefulness of diagnostic tests and screening programs, e.g., colon cancer screening, mammography for breast cancer screening.

Melton LJ: The threat to medical records research. NEJM 1997; 337:1466-1470.
Conclusions

• Research using identifiable medical records has played a vital role in advancing public health and medical knowledge

• There is every indication that it should be able to play as great or greater role in the future, especially as larger records linkage systems are put into place

• A requirement for individual study-specific informed consent for each medical records study, as advocated by some, would make much health research either invalid or so expensive as to be impossible

• “Limited Use Data Sets” and “Data Use Agreements” as described in the August 2002 Final Rule should help lessen the adverse impact of HIPAA on health research